Procedia Computer Science

Volume 29, 2014, Pages 422–431

ICCS 2014. 14th International Conference on Computational Science

Maximizing the Cumulative Influence through a Social Network when Repeat Activation Exists

Chuan Zhou[1], Peng Zhang[1], Wenyu Zang[2], and Li Guo[1]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
{zhouchuan,zhangpeng,guoli}@iie.ac.cn

[2] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
zangwenyu@software.ict.ac.cn

Abstract

We study the problem of employing social networks for propagate influence when repeated activations are involved. While *influence maximization* has been extensively studied as the fundamental solution, it neglects the reality that a user may purchase a product/service repeatedly, incurring cumulative sales of the product/service. In this paper, we explore a new problem of *cumulative influence maximization* that brings the *influence maximization* a step closer to real-world viral marketing applications. In our problem setting, repeated activation exists and we aim to find a set of initial users, through which the maximal cumulative influence can be stimulated in a given time period. To describe the repeated activation behavior, we adopt the voter model to reflect the variation of activations over time. Under the voter model, we formulate the maximization problem and present an effective algorithm. We test and compare our method with heuristic algorithms on real-world data sets. Experimental results demonstrate the utility of the proposed method.

Keywords: Social Networks, Repeated Activations, Voter Model, Cumulative Influence Maximization

1 Introduction

The rapidly increasing popularity of online social networks has created immense opportunities for social network marketing, where the key issue is how to maximize the total sales profit (i.e., maximize the cumulate sales). Viral marketing, among existing solutions, represents one of the most effective ways to promote products and generate more profit on social networks. Viral marketing was first introduced to the data mining community by Domingos and Richardson [10, 22], where cost-effective methods are proposed to promote a new product or technology by giving free or discounted samples to a select group of influential individuals, in the hope that through the word-of-mouth effects over the social network, a large number of product adoptions will occur. Motivated by viral marketing, influence maximization has emerged as a fundamental research problem concerning the propagation of innovations through social networks [6, 7, 8, 13, 16, 17, 18].

Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2014

doi:10.1016/j.procs.2014.05.038

However, existing influence maximization formulation, without considering the reality that a product can be purchased many times in a given time period, cannot precisely describe the cumulative sales (i.e., cumulative influence) maximization problem. Actually, in the literature of influence maximization, it is often assumed that once a user is active, the user will keep the active state forever and cannot be activated again. But the reality is that a user may purchase a product or service repeatedly under social influence. For example, daily necessities, such as toothpastes, makeups and cigarettes, are generally purchased continually by users. Therefore, how to maximize the cumulative sales cannot be simply narrowed down to the influence maximization problem in these scenarios, but demands a new solution framework.

We discuss the problem of *maximizing the cumulative influence* under repeated activations in this paper. Consider a social network as a graph $G = (V, E)$, where V represent users and E represent relationships. Let $S_t \subseteq V$ be the set of nodes who get activated at time t and $|S_t|$ be the size of set S_t. The problem can be described as **how to maximize the cumulative influence $\sum_{t=0}^{T} |S_t|$, when repeated activation exists?** Here, T denotes the time interval that we observe, and similar time constrain can also be found in [5, 20].

The above problem, inspired by influence maximization, can be converted to target an initial subset nodes in the network, such that cumulative influence through the effect of social influence can be maximized. Formally, given a positive integer k, we aim to solve the discrete optimization problem in Eq. (1), where $\mathbb{E}^S$ is the expectation operator with the initial set S.

$$S^* = \arg \max_{|S|=k, S \subseteq V} \mathbb{E}^S \Big[\sum_{t=0}^{T} |S_t| \Big] \tag{1}$$

To solve Eq. (1), the first *challenge* is how to model the variation of S_t when repeated activations happen. Obviously, existing popular propagation models such as Independent Cascade (IC) and Linear Threshold (LT) models [16] are incapable of describing the repeated activation phenomenon, as they mainly describe the range of influence without considering repeated activations.

The second *challenge* is how to develop an efficient algorithm for the optimization problem in Eq. (1). We all know that the influence maximization problem is NP-hard [16] under both IC and LT models, where a line of greedy or heuristic algorithms are proposed. In the same token, we need to answer the question whether the cumulative influence maximization is computationally tractable or an NP-hard problem?

To overcome these challenges, we *first* adopt the framework of the voter model to simulate the diffusions with repeated activation involved in social networks. The voter model, which was first proposed in [9], is probably one of the most basic and natural probability models to represent opinion diffusions in which people may switch opinions back and forth from time to time due to the interactions with other people in the network. The voter model assumes that a node can be activated repeatedly by its newly activated parents and the propagation probability depends on its newly active parents. *Second*, we explore cumulative influence maximization problem from the perspective of stochastic processes and present a simplified calculation procedure for the cumulative influence under the voter model. Furthermore, based on the analysis we propose an *ExactSolution* algorithm for the cumulative influence maximization problem (1), which seeks initial potential nodes with acceptable time cost by experimental results.

The *contribution* of the paper are summarized as follows:

1. We present the cumulative influence maximization problem in social networks, which aims to find the most potential nodes in case of repeated purchase behavior exists. (Section 2)
2. We theoretically analyze the cumulative influence maximization problem and present an efficient solution under the voter model. (Section 3,4)

3. We conduct experiments on real-world data sets to demonstrate the superiority of our new algorithm to other heuristic algorithms. (Section 5)

2 Problem Formulation

We first introduce a variant of the common *voter model* from the single-item-based view, which is slightly different from these in the work [11, 19].

Consider a directed graph $G = (V, E)$ with self-loop and edge label weight $w : E \to [0, 1]$. For convenience, let $w(u, v) = 0$ if $(u, v) \notin E$. For $v \in V$, the set of parents of v is denoted as $Par(v) := \{u \in V,\ (u, v) \in E\}$. For weight w, we assume that for each $v \in V$,

$$\sum_{u \in Par(v)} w(u, v) \leq 1. \tag{2}$$

Given a seed set $S \subseteq V$, the voter model works as follows. Let $S_t \subseteq V$ be the set of nodes that are activated at step $t \geq 0$ with $S_0 = S$. At step $t+1$, every node $v \in V$ can be activated by its newly activated neighbors with probability $\sum_{u \in Par(v) \cap S_t} w(u, v)$. If v is activated successfully, then it is put into the set S_{t+1}. The process ends at a step τ with $S_\tau = \varnothing$. For mathematical tractability, we still denote $S_t = \varnothing$ for $t > \tau$. Obviously the process $(S_t)_{t \geq 0}$ is Markovian because it has no memory.

Given a nonnegative integer T and the initial seed set S, the *T-steps cumulative influence* triggered by S, which is the expected value of total activation numbers from all individuals at time stamps $0, 1, \cdots, T$, is denoted as $\sigma_V^T(S)$, *i.e.*

$$\sigma_V^T(S) := \mathbb{E}^S\Big[\sum_{t=0}^{T} \big|S_t\big|\Big] \tag{3}$$

where the subscript V denotes the voter model used to govern the dynamics of S_t, the superscript T denotes the time span $\{0, 1, \cdots, T\}$, and $\mathbb{E}^S$ is the expectation operator with initial set S. Note that the introduction of time constraint T is due to two reasons. First, the infinite series $\sum_{t=0}^{\infty} \big|S_t\big|$ may be divergent and complicate the solution. Second, many network marketings tend to maximize the cumulative sales in a short amount of time. For example, to advertise icecreams, the marketing time would be mainly restricted in summer.

Let's consider an extreme case where $T = \infty$ in the above definition, then the *long-term cumulative influence* triggered by S, i.e., the expected value of the total activation without time constraint, can be denoted as $\sigma_V(S)$,

$$\sigma_V(S) := \mathbb{E}^S\Big[\sum_{t=0}^{\infty} \big|S_t\big|\Big] = \sigma_V^\infty(S).$$

Based on the above equation, an following question arise: under which condition, $\sum_{t=0}^{\infty} \big|S_t\big|$ is convergent or $\sigma_V(S)$ is a finite value? We will answer this question in the next section.

For a fixed nonnegative integer T (including $T = \infty$), the *cumulative influence maximization problem* under the voter model is to find a subset $S^* \subseteq V$ such that $|S^*| = k$ and $\sigma_V^T(S^*) = \max\big\{\sigma_V^T(S) \bigm| |S| = k, S \subseteq V\big\}$, *i.e.*,

$$S^* = \arg \max_{|S|=k, S \subseteq V} \sigma_V^T(S) \tag{4}$$

where k is a given parameter.

3 Analysis

In this part we analyze the cumulative influence maximization problem under the voter model. We aim to prove that the calculation of cumulative influence $\sigma_V^T(S)$ or $\sigma_V(S)$ under the voter model is tractable, with explicit solutions given in Eqs. (9) and (11). This is very different from the calculation of spread $\sigma_I(S)$ (or $\sigma_L(S)$) under the IC model (or LT model), where the problem is $\#P$-hard [6, 8].

We first introduce two propositions, and then derive the ultimate results in Theorem 1 and Corollary 1. Let $\mathbb{P}^S(v \in S_t)$ denote the probability that a node v gets activated at step t under the seed set S in the voter model, then we have the first proposition as follows,

Proposition 1. *For $S \subseteq V$ and $T \geq 0$, the cumulative influence $\sigma_V^T(S)$ under the voter model can be resolved as*

$$\sigma_V^T(S) = \sum_{t=0}^{T} \sum_{v \in V} \mathbb{P}^S(v \in S_t). \tag{5}$$

Proof: From the definition of cumulative influence in Eq. (3), we have

$$\begin{aligned} \sigma_V^T(S) &= \sum_{t=0}^{T} \mathbb{E}^S\Big[\big|S_t\big|\Big] = \sum_{t=0}^{T} \mathbb{E}^S\Big[\sum_{v \in V} I_{S_t}(v)\Big] \\ &= \sum_{t=0}^{T} \sum_{v \in V} \mathbb{E}^S\big[I_{S_t}(v)\big] = \sum_{t=0}^{T} \sum_{v \in V} \mathbb{P}^S(v \in S_t) \end{aligned}$$

where $I_{S_t}(v)$ is an indicative function, if $v \in S_t$, then $I_{S_t}(v) = 1$; otherwise, $I_{S_t}(v) = 0$. □

Proposition 1 reveals that we can treat the cumulative influence measure $\sigma_V^T(S)$ as a summation of all T propagation steps of local probabilities $\{\mathbb{P}^S(v \in S_t) : t \geq 0, v \in V\}$. Based on Proposition 1, a following question is, *what is the relationship between two adjacent sets $\{\mathbb{P}^S(v \in S_t) : v \in V\}$ and $\{\mathbb{P}^S(v \in S_{t-1}) : v \in V\}$?*

Proposition 2. *For each $t \geq 1$, we have the following equation*

$$\mathbb{P}^S(v \in S_t) = \sum_{u \in V} \mathbb{P}^S\big(u \in S_{t-1}\big) w(u, v). \tag{6}$$

Proof: For $t \geq 1$, by the definition of conditional expectation and the voter model,

$$\mathbb{P}^S(v \in S_t) = \mathbb{E}^S\big[\mathbb{P}^S(v \in S_t | S_0, \cdots, S_{t-1})\big] = \mathbb{E}^S\Big[\big(\sum_{u \in S_{t-1}} w(u, v)\big)\Big]$$

$$= \mathbb{E}^S\Big[\big(\sum_{u \in V} I_{\{u \in S_{t-1}\}} w(u, v)\big)\Big] = \sum_{u \in V} \mathbb{P}^S\big(u \in S_{t-1}\big) w(u, v).$$

In the above derivation, the first '=' stems from the conditional expectation, the second one is from the definition of the voter model. □

Proposition 2 implies that the activated probabilities at different time steps share some iterative relationship, which provides a tractable basis for further investigation. For simplicity, we rewrite the results in Propositions 1 and 2 using matrix notations. For $t \geq 0$, denote the row vector $\Pi_t^S = \big(\pi_t^S(v)\big)$ as the probabilities of nodes being activated at step t, *i.e.*, $\pi_t^S(v) := \mathbb{P}^S(v \in S_t)$. Then, Proposition 1 can be rewritten as

$$\sigma_V^T(S) = \sum_{t=0}^{T} \Pi_t^S \cdot \mathbf{1} \tag{7}$$

where $\mathbf{1}$ is a column vector with all elements being 1, and Proposition 2 can be rewritten as

$$\Pi_t^S = \Pi_{t-1}^S \cdot W \tag{8}$$

where W is the weight matrix with the (u, v) position's element being $w(u, v)$. Then we have Theorem 1 as below,

Theorem 1. *The cumulative influence function $\sigma_V^T(S)$ can be represented as a summation of influence at different time stamps, i.e.,*

$$\sigma_V^T(S) = \sum_{t=0}^{T} \Pi_0^S \cdot W^t \cdot \mathbf{1}. \tag{9}$$

Proof: By the iteration in Eq.(8), we can easily get that $\Pi_t^S = \Pi_0^S \cdot W^t$, where $\Pi_0^S = \big(\pi_0^S(v)\big) = \big(I_S(v)\big)$. Incorporating this result into Eq. (7), we can obtain Eq. (9). Note that $\Pi_0^S \cdot W^t \cdot \mathbf{1}$ is a real number after matrix calculation. □

Based on Eq.(9) in Theorem 1, one may easily raise two questions,

- *The function $\sigma_V^T(S)$ is resolved into a summation of series $\sum_{t=0}^{T} \Pi_0^S \cdot W^t \cdot \mathbf{1}$, if we let $T = \infty$ to get $\sigma_V(S) = \sum_{t=0}^{\infty} \Pi_0^S \cdot W^t \cdot \mathbf{1}$, then in what condition the series will be convergent?*
- *If the relaxed series is convergent, what's the convergence limit, i.e., $lim_{n\to\infty} \sum_{t=0}^{\infty} \Pi_0^S \cdot W^t \cdot \mathbf{1}$ =?.*

In the sequel, we derive Corollary 1 to answer above two questions.

Corollary 1. *If the weight satisfies the condition (which is a further requirement to Eq. (2))*

$$\max_v \sum_{u\in V} w(u,v) < 1, \tag{10}$$

then the series $\sum_{t=0}^{\infty} \Pi_0^S \cdot W^t \cdot \mathbf{1}$ is convergent, and the limit of convergence exists as,

$$\sigma_V(S) = \lim_{T\to\infty} \sigma_V^T(S) = \Pi_0^S \cdot (E - W)^{-1} \cdot \mathbf{1} \tag{11}$$

where E is a unit matrix and $(E-W)^{-1}$ is the inverse of $(E-W)$. Besides, we also have

$$\sigma_V(S) - \sigma_V^T(S) \leq \frac{N \cdot \parallel W \parallel_1^{T+1}}{1 - \parallel W \parallel_1} \tag{12}$$

where the 1-norm of matrix W is defined as $\parallel W \parallel_1 := \max_v \sum_{u\in V} w(u,v)$ and N is the size of the graph.

Proof: By matrix analysis [15], condition (10) implies $\parallel W \parallel_1 < 1$, which ensures the convergence of matrix series $\sum_{t=0}^{\infty} W^t$ with the limit $(E-W)^{-1}$. Hence we have $\sigma_V(S) = \Pi_0^S \cdot \Big(\sum_{t=0}^{\infty} W^t\Big) \cdot \mathbf{1} = \Pi_0^S \cdot (E-W)^{-1} \cdot \mathbf{1}$, where the first '=' is from Theorem 1. As to Eq. (12), it is derived like this

$$\begin{aligned} \sigma_V(S) - \sigma_V^T(S) &= \Pi_0^S \cdot \Big((E-W)^{-1} - \sum_{t=0}^{T} W^t\Big) \cdot \mathbf{1} \\ \leq &\parallel \Pi_0^S \parallel_1 \cdot \parallel (E-W)^{-1} - \sum_{t=0}^{T} W^t \parallel_1 \cdot \parallel \mathbf{1} \parallel_1 \leq \frac{N \cdot \parallel W \parallel_1^{T+1}}{1 - \parallel W \parallel_1} \end{aligned}$$

In the above derivation, the first '$\leq$' is from the submultiplicative of matrix norm (*i.e.*, $\parallel A \cdot B \parallel \leq \parallel A \parallel \cdot \parallel B \parallel$). The second '$\leq$' is due to $\parallel \Pi_0^S \parallel_1 = 1$, $\parallel \mathbf{1} \parallel_1 = \mathrm{N}$, and $\parallel (E-W)^{-1} - \sum_{t=0}^{T} W^t \parallel_1 \leq \frac{\|W\|_1^{T+1}}{1-\|W\|_1}$ in matrix norm theory. □

Algorithm 1: ExactSolution

1: Input: weight matrix W of a graph $G = (V, E)$, budget k, and time constraint T
2: Output: k nodes set S^* with the maximum T-steps cumulative influnece
3: **Score** $\leftarrow \sum_{t=0}^{T} W^t \cdot \mathbf{1}$ $\big($*or* $(E - W)^{-1} \cdot \mathbf{1}$ if $T = \infty\big)$
4: Select the biggest k nodes in **Score** as the output S^*

4 Solution

Based on the analysis above, we provide an efficient algorithm to the cumulative influence maximization problem. Inspired by Eqs. (9) and (11), in order to maximize the T*-steps cumulative influence* or the *long-term cumulative influence*, we need to choose an initial seed set S^* with the highest values in the column vector $\sum_{t=0}^{T} W^t \cdot \mathbf{1}$ or $(E - W)^{-1} \cdot \mathbf{1}$. We call it the **ExactSolution** and summarize below.

Although ExactSolution algorithm presents an optimal solution, the matrix operations involved will get expensive in the respect of running time and memory, especially the inverse $(E - W)^{-1}$ will become intractable, when the size of network is enormous. To overcome this difficulty, we adopt the following method to calculate $\sum_{t=0}^{T} W^t \cdot \mathbf{1}$ and $(E-W)^{-1} \cdot \mathbf{1}$. For $t \geq 0$, we denote the column vector $\mathbf{a}_t := W^t \cdot \mathbf{1}$, so we obtain

$$\sum_{t=0}^{T} W^t \cdot \mathbf{1} = \sum_{t=0}^{T} \mathbf{a}_t, \tag{13}$$

and

$$(E - W)^{-1} \cdot \mathbf{1} = \sum_{t=0}^{\infty} W^t \cdot \mathbf{1} = \sum_{t=0}^{\infty} \mathbf{a}_t, \tag{14}$$

where $\mathbf{a}_{t+1} = W \cdot \mathbf{a}_t$. By using the iteration, to acquire $(E - W)^{-1} \cdot \mathbf{1}$, we only need to sum up $\mathbf{a}_0, \mathbf{a}_1, \cdots$ until some $\mathbf{a}_n$ with L_1-norm less than 10^{-3}. This transformation saves memory space during calculation, as it stores vectors instead of matrices. Besides, an alternative method to derive $(E - W)^{-1} \cdot \mathbf{1}$ is to solve the linear equations $(E - W) \cdot \mathbf{Score} = \mathbf{1}$. Nevertheless, it is still necessary to take heuristic algorithms into comparison in experiments.

5 Experiments

We conduct experiments on four real-world data sets to evaluate the cumulative influence and the ExactSolution algorithm. We implement the algorithm in C++ with the Standard Template Library (STL). All experiments are run on a Linux (Ubuntu 11.10) machine with six-core 1400 MHz AMD CPU and 32 GB memory.

Table 1: Four real-world networks.

Dataset	Facebook	Digger	Twitter	Epinions
#Node	4093	8,193	32985	51782
#Edge	176468	56,440	763713	476491
Average Degree	21.6	6.9	23.2	9.2
Maximal Degree	1045	850	674	190

We use four real-world data sets, summarized in Table 1, for testing and comparisons. The Facebook and Twitter data are obtained from *http://snap.stanford.edu/data/*. The Digger data is available at *http://arnetminer.org/heterinf*. The Epinions data, a who-trust-whom online social network, can be obtained from *http://snap.stanford.edu/data/*. The details of the data sets are listed in Table 1, where degree means out-degree.

We implement four benchmark algorithms, DEGREE [16], PAGERANK [4], DEGREEDISCOUNT [7], and RANDOM for comparisons.

- DEGREE[16]. A heuristic algorithm based on "degree centrality", with high-degree nodes as influential ones. The seeds are the nodes with the k highest out-degrees.
- PAGERANK [4]. A link analysis algorithm which ranks the importance of pages in a Web graph. We implement with a damping factor of 0.85 and pick the k highest-ranked nodes as seeds.
- DEGREEDISCOUNT [7]. A degree discount heuristic algorithm which deducts the overlap effect in DEGREE algorithm.
- RANDOM. It simply selects k random vertices in the graph as the seed set, which is also evaluated in [16].

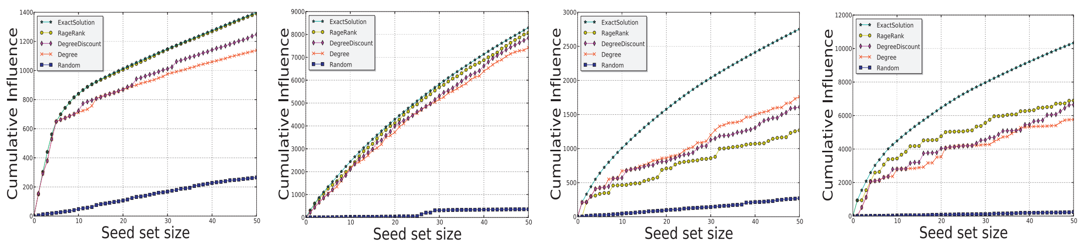
We measure the comparisons in two aspects: quality of the seed set (*i.e.* cumulative influence) and efficiency of the algorithm (*i.e.* running time). Moreover, to obtain the cumulative influnece of benchmark methods for each seed set, we substitute them into Eq. (9) or Eq. (11). For ease of comparisons, we ignore other centrality measures, such as the distance centrality and betweenness centrality as heuristics, due to their inefficiency in the running time [7].

We assign the propagation probability of each directed link $(u, v) \in E$ in the network for the voter model as follows,

$$w(u,v) = \begin{cases} \alpha & if \quad u = v \\ \frac{1-\alpha}{|Par(v)|} & if \quad u \in Par(v) \backslash \{v\} \end{cases}$$

where $|Par(v)|$ denotes the number of parents of a node v. We set $\alpha = 0.5$. In our experiments, an undirected graph can be also regarded as a bidirectional graph.

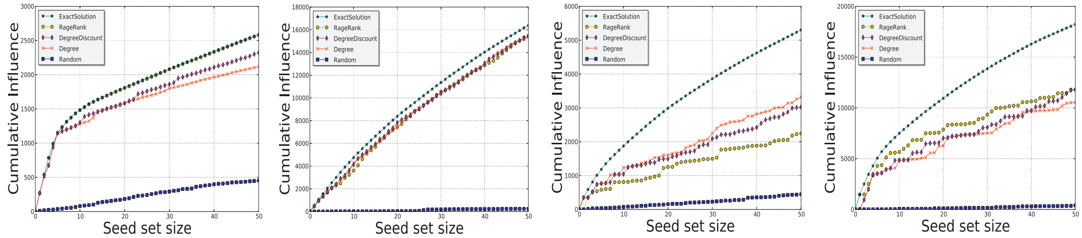
Cumulative Influence $\sigma_V^T(S)$ is an important measure for comparisons. In this part, we run tests on the four data sets to obtain cumulative influence results *w.r.t.* parameter k (the seed set size), where k increases from 1 to 50. We list the results in Fig. 1 and Fig. 2 with $T = 5$ and $T = \infty$ respectively. For easy reading, in the figures, the legend ranks the algorithms top-down based on their average cumulative sales values. From the results, we can observe that no benchmark algorithms can match with the ExactSolution algorithm on the four data sets. In contrast, the PageRank algorithm provides the nearest approximation in most cases, except on Twitter data. The results from Random show that maximizing the cumulative influence is far from trivial.



(a) Facebook (b) Digger (c) Twitter (d) Epinions

Figure 1: Cumulative influence *w.r.t.* seed size k on the four data sets, $T = 5$.

Time Cost of selecting 50 seeds is shown in Figure 3, where ExactSolution_5 means $T = 5$ and ExactSolution_∞ means $T = \infty$. From the results, we can observe that the heuristic algorithms, Random, Degree and DegreeDiscount, are very fast in selecting candidate nodes (which takes less than 1 second). The PageRank algorithm is slightly slower than the above



(a) Facebook (b) Digger (c) Twitter (d) Epinions

Figure 2: Cumulative influence *w.r.t.* seed size k on the four data sets, $T = \infty$.

three algorithms due to heavy iterations. The ExactSolution algorithm takes the longest time, but still acceptable (in a minute level). By comparing the runtime of ExactSolution_5 and ExactSolution_∞, we observe that the latter is about $4-6$ times slower than the former one.

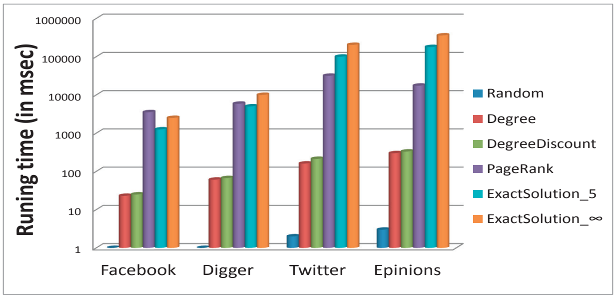


Figure 3: Runtime comparison.

6 Related Work

Influence maximization. Influence maximization is defined as finding a small subset of k nodes that maximizes spread of influence $\sigma_M(S)$ in social networks, based on a given stochastic influence propagation model M. In this context, the *Independent Cascade* (IC) and *Linear Threshold* (LT) models [16] are two popularly used models. Influence maximization under both IC and LT models is NP-hard, and the spread function is monotone and submodular [16]. A set function $f: 2^V \to \mathbb{R}^+$ is monotone, if $f(S) \leq f(T)$ whenever $S \subseteq T \subseteq V$. The set function is submodular, if $f(S \cup \{w\}) - f(S) \geq f(T \cup \{w\}) - f(T)$ for all $S \subseteq T$ and $w \in V \backslash T$.

Exploiting these two properties, Kempe et al. [16] presented a simple greedy algorithm which repeatedly chooses the node with the maximum marginal gain and adds it to the seed set, until the budget k is reached. The simple greedy algorithm can approximate the solution within a factor of $(1 - 1/e - \epsilon)$ for any $\epsilon > 0$.

Unfortunately, the simple greedy algorithm suffers from great inefficiency. Hence considerable work has been conducted to tackle this limitation. One direction is to propose heuristic solutions, *e.g.*, DegreeDiscount [7] and ShortestPath [17]. These heuristic algorithms can reduce computational cost in orders of magnitude, with competitive results of the influence spread level. However, none of them has a theoretical guarantee on the reliability of the results. The other direction is to improve the original greedy algorithm. A representative work, by Leskovec et al. [18], exploited the submodular property of the objective function, and proposed a Cost-Effective Lazy Forward selection (CELF) algorithm, which improves the running time of the simple greedy algorithm by up to 700 times. Following the same logic, Goyal et al. proposed CELF++ [12], an extension of CELF, that further reduces the number of spread estimation calls, leading to $35\% - 55\%$ faster than CELF. Recently, Zhou et al. in [23, 24] further enhance the CELF by the upper bound based approach UBLF, in their method the Monte-Carlo calls in the first round are drastically reduced comparing with the CELF.

Viral marketing on social networks. There have been several efforts to monetize social networks recently, and maximizing product adoption or profit is a promising attempt along this line. Bhagat et al. [2] addressed the product adoption maximization by differentiating between product adoption and influence in their LT-C model. In [21], the authors extended the classical Linear Threshold (LT) model to incorporate prices and valuations, and factor them into users' decision-making process of adopting a product. Considerable work has also been done on pricing in social networks. One related problem is that of revenue maximization, introduced by Hartline et al. [14]. They studied optimal marketing for digital goods in social networks and proposed the influence-and-exploit (IE) framework. In IE, seeds are offered free samples, and the seller can approach other users in a random sequence, bypassing the network structure. Arthur et al. [1] adopted IE to study a similar problem in which users arrive in a sequence decided by a cascade model (IC). Bloch et al. [3] formulated pricing in social networks as simultaneous-move games and studied equilibria of the games. However, none of above works take the repeated purchase behavior into account to increase the gain.

7 Conclusions

In this paper we propose the cumulative influence maximization problem by considering the scenario of repeated activations in social networks. We present its motivations, discuss its difference from the classical influence maximization problem, explore its unique properties by stochastic analysis, and present an exact algorithm as the solution under the voter model. Also we show the utility of the proposed ExactSolution algorithm by experiments.

There are several interesting future directions. First, the discrete formulation of propagation time can be further modified to a tractable continuous-time version; second, we regard the network weight as predefined in the voter model, actually how to infer these parameters from real-world data is also challenging.

Acknowledgments

This work was supported by the NSFC (No. 61370025), 863 projects (No. 2011AA010703 and 2012AA012502), 973 project (No. 2013CB329605 and 2013CB329606), and the Strategic Leading Science and Technology Projects of Chinese Academy of Sciences(No.XDA06030200).

References

[1] David Arthur, Rajeev Motwani, Aneesh Sharma, and Ying Xu. Pricing strategies for viral marketing on social networks. In *Internet and Network Economics*, pages 101–112. Springer, 2009.

[2] Smriti Bhagat, Amit Goyal, and Laks VS Lakshmanan. Maximizing product adoption in social networks. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 603–612. ACM, 2012.

[3] Francis Bloch, Nicolas Quérou, et al. Pricing in networks. 2008.

[4] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1):107–117, 1998.

[5] Wei Chen, Wei Lu, and Ning Zhang. Time-critical influence maximization in social networks with time-delayed diffusion process. In *AAAI*, 2012.

[6] Wei Chen, Chi Wang, and Yajun Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1029–1038. ACM, 2010.

[7] Wei Chen, Yajun Wang, and Siyu Yang. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 199–208. ACM, 2009.

[8] Wei Chen, Yifei Yuan, and Li Zhang. Scalable influence maximization in social networks under the linear threshold model. In *IEEE 10th International Conference on Data Mining (ICDM)*, pages 88–97. IEEE, 2010.

[9] Peter Clifford and Aidan Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581–588, 1973.

[10] Pedro Domingos and Matt Richardson. Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 57–66. ACM, 2001.

[11] Eyal Even-Dar and Asaf Shapira. A note on maximizing the spread of influence in social networks. In *Internet and Network Economics*, pages 281–286. Springer, 2007.

[12] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. Celf++: optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th international conference companion on World wide web*, pages 47–48. ACM, 2011.

[13] Jing Guo, Peng Zhang, Chuan Zhou, Yanan Cao, and Li Guo. Personalized influence maximization on social networks. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 199–208. ACM, 2013.

[14] Jason Hartline, Vahab Mirrokni, and Mukund Sundararajan. Optimal marketing strategies over social networks. In *Proceedings of the 17th international conference on World Wide Web*, pages 189–198. ACM, 2008.

[15] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[16] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

[17] Masahiro Kimura and Kazumi Saito. Tractable models for information diffusion in social networks. In *Knowledge Discovery in Databases: PKDD 2006*, pages 259–271. Springer, 2006.

[18] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 420–429. ACM, 2007.

[19] Yanhua Li, Wei Chen, Yajun Wang, and Zhi-Li Zhang. Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In *WSDM*, 2013.

[20] Bo Liu, Gao Cong, Yifeng Zeng, Dong Xu, and Y Chee. Influence spreading path and its application to the time constrained social influence maximization problem and beyond. *IEEE Transactions on Knowledge and Data Engineering*, 2013.

[21] Wei Lu and Laks VS Lakshmanan. Profit maximization over social networks. In *IEEE 12th International Conference on Data Mining (ICDM)*, pages 479–488. IEEE, 2012.

[22] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 61–70. ACM, 2002.

[23] Chuan Zhou, Peng Zhang, Jing Guo, and Li Guo. An upper bound based greedy algorithm for mining top-k influential nodes in social networks. In *Proceedings of the 23th international conference companion on World wide web*. ACM, 2014.

[24] Chuan Zhou, Peng Zhang, Jing Guo, Xingquan Zhu, and Li Guo. Ublf: An upper bound based approach to discover influential nodes in social networks. In *IEEE 13th International Conference on Data Mining (ICDM)*, pages 907–916. IEEE, 2013.